Performance-Complexity Analysis for MAC ML-based Decoding with User Selection

Hsiao-feng (Francis) Lu, Petros Elia, and Arun Singh

Abstract—This work explores the rate-reliability-complexity limits of the quasi-static K-user *multiple access channel* (MAC), with or without feedback. Using high-SNR asymptotics, the work first derives bounds on the computational resources required to achieve near-optimal (ML-based) decoding performance. It then bounds the (reduced) complexity needed to achieve any (including suboptimal) diversity-multiplexing performance tradeoff (DMT) performance, and finally bounds the same complexity, in the presence of feedback-aided user selection. This latter effort reveals the ability of a few bits of feedback not only to improve performance, but also to reduce complexity. In this context, the analysis reveals the interesting finding that proper calibration of user selection can allow for near-optimal ML-based decoding, with complexity that need not scale exponentially in the total number of codeword bits. The derived bounds constitute the best known performance-vs-complexity behavior to date for ML-based MAC decoding, as well as a first exploration of the complexity-feedback-performance interdependencies in multiuser settings.

I. INTRODUCTION

A. Multiple access system model

We consider a symmetric multiple access channel (MAC) with K single-antenna users, communicating to a receiver with n_r receiving antennas, over a quasi-static fading channel. Each user i, $i = 1, 2, \ldots, K$ communicates over the same duration of T time slots, while the receiver accumulates an $n_r \times T$ signal matrix $\boldsymbol{Y}$ taking the form

$$\boldsymbol{Y} = \sqrt{\text{SNR}} \sum_{i=1}^{K} \boldsymbol{h}_i \, \boldsymbol{x}_i^\top + \boldsymbol{W} \tag{1}$$

where $\boldsymbol{h}_i \sim \mathbb{CN}\left(\boldsymbol{0}, \boldsymbol{I}_{n_r}\right)$ is the ith user channel vector with independent identically distributed (i.i.d.) zero-mean unit-variance $\mathbb{CN}\left(0, 1\right)$ complex Gaussian entries, where $\boldsymbol{W}$ is the received noise matrix with i.i.d. $\mathbb{CN}\left(0, 1\right)$ entries, where SNR denotes the signal to noise ratio, and where $\boldsymbol{x}_i = [x_{i,1} \cdots x_{i,T}]^\top$ is a scaled version of a T-length code vector of user i satisfying an average power constraint $\mathbb{E}\left\|\boldsymbol{x}_i\right\|^2 \leq T$, $i = 1, \ldots, K$.

1) Exploring the scenario of outage-limited communications with bounded computational resources: We here consider the outage-limited MAC setting, where the channel in (1) is randomly drawn but it remains fixed throughout the coding duration of T channel uses. This common assumption corresponds to scenarios where the channel changes slowly, and where communication takes place under strict latency

constraints that do not allow for encoding over a large number of fading realizations.

This same outage-limited setting can often experience reduced reliability due to the event of outage where the instantaneous channel cannot support the user rates, and due to the fact that in cases like the MAC, the maximum allowable rates are diminished by interference. Moreover, to meet the strong demands for faster data rates on wireless and cellular channels, most communications systems would opt to operate closer and closer to the theoretical limits of capacity, at the further expense of reliability. This in turn, naturally forces the need for high-performance transceivers that will not further sacrifice reliability performance, but which can often be computationally expensive, or even computationally prohibitive.

What happens though to this performance if communication takes place under strict constraints on the computational resources that can be used to encode and decode? Equivalently, one can ask, what computational resources are needed to achieve a certain rate-reliability performance. The main objective of our work is to provide some understanding of such complexity-performance interdependencies that are crucial in the MAC.

2) Complexity-performance interdependencies in the MAC.: Naturally there are many such interdependencies between the key parameters of SNR, rate, computational complexity and reliability. For example it is easy to imagine that, typically, decreasing computational resources can potentially increase the probability of error by placing a constraint on the coding duration T, by forcing the use of less complex transceivers with suboptimal performance, or even by requiring that the decoding effort be terminated early (computational halt) at the risk of additional errors. Similarly, increased user rates can mean larger and denser codes, with more errors and larger decoding algorithmic efforts. Along the same lines, fixing the rate and increasing the SNR, will make the codewords sparser and thus possibly easier to differentiate and decode.

B. Complexity-performance measures and high-SNR approximations

Let $N_{\max}$ denote the amount of computational reserves, in floating point operations (flops) per T channel uses, that the system is endowed with, in the sense that after $N_{\max}$ flops, the decoder must simply terminate, potentially prematurely and before completion of the task, thus declaring an error. Also let P_e be the probability of error associated to the decoder, in the presence of the aforementioned computational constraints. Motivated by the need for reduced-complexity high-performance ML-based decoders (cf. [7]–[11]), we explore

the properties of this quadruple $\big(\mathrm{SNR}, R, P_e, N_{\max}\big)$, for ML-based MAC decoding.

1) High SNR analysis of performance and complexity:
Towards making sense of the complexity-vs-performance interdependencies, we will exploit carefully-chosen asymptotic bounds, where analytical tools can rigorously and concisely approximate these interdependencies, allowing for cleaner insights that can yield impact. Such asymptotic bounds (generally mapping a random problem to its near-deterministic limits), help to identify where the largest gaps in our understanding may lie. We will here focus on *high-SNR* asymptotic bounds but note that the derived approximations can offer insights in operational regimes of moderate-to-high SNR values. Such emphasis on the moderate-to-high SNR values, can better capture the core of the complexity-vs-performance problem because such operational regions can eventually support higher rates, corresponding to larger codebooks that are typically — but not always, as we will surprisingly see later on — harder to decode.

Towards this, we let each user's rate scale as $R = r \log_2 \mathrm{SNR}$, where r denotes the *multiplexing gain* [6] describing the density of the constellation, and serving as a measure of how far each user's rate is from the capacity of an AWGN channel. Each user i employs a code $\mathcal{X}_{r,i}$ of cardinality $|\mathcal{X}_{r,i}| = 2^{RT} = \mathrm{SNR}^{rT}$, and the receiver uses a decoder $\mathcal{D}_r$, which is considered to make an error if any of the K users' messages is not decoded correctly. We restrict our attention to the class of lattice codes and joint ML decoders, which we describe in Section I-D that follows.

2) Complexity and reliability exponents: For complexity analysis we adopt the high-SNR approach in [1], [2], where for some r, for some encoders $\{\mathcal{X}_{r,i}\}_{i=1}^{K}$, and for a decoder $\mathcal{D}_r$, the *complexity exponent* $c(r)$ was defined to be

$$c(r) := \lim_{\mathrm{SNR} \to \infty} \frac{\log N_{\max}(r)}{\log \mathrm{SNR}}. \tag{2}$$

We also adopt the well-known DMT approach of [6] where, in the same high SNR regime, the *diversity gain* $d(r)$ takes the form

$$d(r) := -\lim_{\mathrm{SNR} \to \infty} \frac{\log P_e(r)}{\log \mathrm{SNR}}. \tag{3}$$

Using this asymptotic approach, the work in [5] has shown that the optimal K-user MAC DMT performance, takes the form

$$d_{\mathrm{mac}}^*(r) := \begin{cases} n_r(1-r), & \text{if } 0 \le r \le \frac{n_r}{K+1}, \\ d_{K,n_r}^*(Kr), & \text{if } \frac{n_r}{K+1} < r \le \frac{n_r}{K}, \end{cases} \tag{4}$$

where $d_{m,n}^*(r)$ denotes the optimal DMT of an $(m \times n)$ MIMO channel (cf. [6] for its exact characterization). Our first effort here will be to bound the complexity exponent $c(r)$ that can guarantee any MAC-DMT performance $d(r) \le d_{\mathrm{mac}}^*(r)$. Before we do that, let us try to get some insight on these two competing exponents.

3) Insight on $d(r)$ and $c(r)$: To gain some insight, we note that in terms of the MAC-DMT performance, the above expression in (4) reflects the existence of two distinct r regions; the *lightly-loaded* multiplexing gain region $0 \le r \le \frac{n_r}{K+1}$ where the MAC exhibits single-user behavior as if there

were no multiuser interference, and the *heavily-loaded* region $\frac{n_r}{K+1} < r \le \frac{n_r}{K}$, where the MAC exhibits an antenna-pooling behavior.

In terms of the complexity exponent $c(r)$, it is easy to see that in this MAC setting, there is a total of $2^{RKT} = \mathrm{SNR}^{rKT}$ codewords (jointly from all users), which means that a *brute-force* joint ML decoder would always make 2^{RKT} codeword visits [1]. Thus in terms of exponents, this means that such a brute-force optimal ML decoder, can achieve the optimal $d_{\mathrm{mac}}^*(r)$ with a required complexity exponent $c(r) = rKT$ that is indeed the maximum (meaningful) complexity exponent in this setting[2]. We will show that a properly designed sphere decoder can achieve this same ML performance, with a reduced complexity exponent $c(r) < rKT$. To give the reader an idea, we can recall from the work in [1] that, if for example, there is only one user ($K = 1$) who has n_t transmit antennas, and the receiver has $n_r \ge n_t$ antennas, then — again in the high SNR limit — this same ML performance can be achieved with computational resources that are substantially smaller than brute force, and which corresponded to an (effective) complexity exponent that was a non-monotonic[3] piece-wise linear function in r, of the form

$$c(r) = rT\left(1 - \frac{r}{n_t}\right), \quad r = 0, 1, 2, \ldots, n_t.$$

We will do something similar here, for the multiuser case $K \ge 1$, and we will also explore the effect of a few bits of feedback on the $d(r)$ and $c(r)$.

C. Notation and assumptions

Following [6], we use $\doteq$ to denote the *exponential equality*, i.e., a function $f(\mathrm{SNR})$ is said to be $f(\mathrm{SNR}) \doteq \mathrm{SNR}^b$ if and only if $\lim_{\mathrm{SNR} \to \infty} \frac{\log f(\mathrm{SNR})}{\log \mathrm{SNR}} = b$. Exponential inequalities such as $\dot{\le}, \dot{\ge}$ are similarly defined. By $s = \lceil x \rceil$ we mean the smallest integer $s \ge x$, and by $t = \lfloor x \rfloor$ we mean the largest integer $t \le x$. Capital boldface letters are reserved for matrices, and the lower-case boldface ones are for column vectors. $\mathbf{A}^\dagger$ is the Hermitian transpose of matrix $\mathbf{A}$, and $(x)^+ := \max\{x, 0\}$. Finally, we define $\nu := \min\{K, n_r\}$.

In terms of assumptions, we consider fading coefficients that have a circularly symmetric complex-normal distribution and which are i.i.d. in space. As stated, the coding duration is smaller than the coherence interval of the channel, and hence the fading is randomly drawn, but it is held fixed throughout the communication process. We will assume that the receiver has full channel state information, i.e., knows completely the channel vectors $\mathbf{h}_i$ of every user i, while the users have no knowledge of $\mathbf{h}_i$ in the case of no feedback. Furthermore,

---

[1] The number of codeword visits is — in the high SNR setting — in the same order as the number of computational flops.

[2] Considering decoders with higher complexity than brute-force ML, is unnecessary because ML decoders are already optimal.

[3] This non monotonic expression is maximum somewhere in the mid-ranges of r, suggesting that the computational resources to achieve the asymptotically optimal complexity need not necessarily increase with increasing rate. For example, for the simple case of $n_t = 2$, in the presence of the minimum required $T = n_t = 2$ in order to achieve the optimal DMT of $d_{2,n_r}^*(r)$, the complexity exponent was increasing as $c(r) = r$ for $0 \le r \le 1$, and was then decreasing as $c(r) = 2 - r$ for $1 \le r \le 2$.

the rate of communication R is kept constant and thus does not change as a function of the channel and of the feedback. Finally, different assumptions regarding the structure of the code and decoder, will be presented immediately below in Section I-D. Assumptions on the feedback-aided user-selection algorithm will be presented in Sec. III.

D. Lattice encoders and joint decoder

Directly from (1) we have the joint real-valued vectorized model which takes the form

$$y = \sqrt{\mathrm{SNR}} H x + w \tag{5}$$

where

$$H = I_T \otimes \begin{bmatrix} \mathrm{Re}\{H_{\mathrm{eq}}\} & -\mathrm{Im}\{H_{\mathrm{eq}}\} \\ \mathrm{Im}\{H_{\mathrm{eq}}\} & \mathrm{Re}\{H_{\mathrm{eq}}\} \end{bmatrix} \tag{6}$$

where

$$H_{\mathrm{eq}} = [h_1 \ \cdots \ h_K]$$

represents the entire $(n_r \times K)$ channel fading matrix, where

$$x = [\mathrm{Re}\{x_{1,1}\} \, \mathrm{Im}\{x_{1,1}\} \cdots \mathrm{Re}\{x_{K,1}\} \, \mathrm{Im}\{x_{K,1}\}$$
$$\mathrm{Re}\{x_{1,2}\} \, \mathrm{Im}\{x_{1,2}\} \ \cdots \ \mathrm{Re}\{x_{K,T}\} \, \mathrm{Im}\{x_{K,T}\}]^\top$$

is the joint codeword vector aggregated over all users, and where y and w are defined similarly to x. The joint codeword x is decoded using a joint ML-based decoder which, in its brute-force form, is MAC DMT optimal (cf. [5], [15]), i.e., achieves the optimal MAC DMT $d^*_{\mathrm{mac}}(r)$. x is taken from a (sequence of) full-rate linear (lattice) code(s) $\mathcal{X}_r = \mathcal{X}_{r,1} \oplus \cdots \oplus \mathcal{X}_{r,K}$, where $\mathcal{X}_{r,i} = \Lambda_{r,i} \cap \mathcal{R}_{r,i} \subset \mathbb{R}^{2T}$ is the corresponding lattice code for the ith user consisting of those elements of the rank $2T$ lattice $\Lambda_{r,i}$ that lie inside the *shaping region* $\mathcal{R}_{r,i}$, which is properly chosen to meet the rate requirement $|\mathcal{X}_{r,i}| = 2^{RT}$ as well as the average power constraint. The region $\mathcal{R}_{r,i}$ is a compact convex subset of $\mathbb{R}^{2T}$. Specifically, we set $\Lambda_{r,i} := \mathrm{SNR}^{-\frac{r}{2}} \Lambda_i$, to be a scaled lattice of another lattice Λ_i whose generator matrix is denoted by G_i. Thus for $G = \mathrm{diag}(G_1, \ldots, G_K)$, the overall codeword is given by $x = \mathrm{SNR}^{-\frac{r}{2}} G s$ for some $s \in \mathbb{Z}^{2KT}$. Substituting this into (5) yields the following equivalent channel input-output relation which will be used for sphere decoding of s

$$y = M s + w \tag{7}$$

where $M := \mathrm{SNR}^{\frac{1-r}{2}} H G$.

This joint ML (or lattice) decoder is implemented as a bounded-search sphere decoder (SD). For SD algorithm details, readers are referred to [1], [2], [26] however for clarity of exposition, wherever necessary, essential details are provided during the complexity analysis. Implicit to the use of a sphere decoder is a chosen search radius δ, a chosen decoding order corresponding to an order with which symbols of s are decoded, and a time-out policy that terminates the decoder once the algorithm exceeds a certain computational threshold. The termination policies that we use will be clarified depending on the setting. Our results will optimize over δ by setting $\delta := \sqrt{z \log \mathrm{SNR}}$, for a properly chosen $z > 0$. The idea here is that the search radius should be big enough so

that — loosely speaking — the probability that AWGN noise has a norm larger than this radius, is sufficiently smaller than the probability of error under brute-force ML. At the same time, this radius needs to remain sufficiently small so that the number of elements within the search sphere is small enough most of the time. Finally, the bounds will hold irrespective of the decoding order. As a result, we will henceforth limit reference to the search radius and the decoding order, mainly in proofs. As suggested before, the derivations focus on ML-based decoding, but given the aforementioned performance-and-complexity equivalence between ML and *regularized* lattice based decoding [2], these same results extend automatically to the latter.

We finally note that the validity of the presented bounds depends on the existence of actual coding schemes that meet them, and which will be identified here, together with the associated SD implementation and halting policies. Regarding the codes, we hasten here to say that all rate-reliability results can be achieved with uncoded QAM transmission for any n_r. This is one of the crucial contributions of this paper, and it is presented in Theorem 2.

II. PERFORMANCE-COMPLEXITY TRADEOFF FOR MAC ML DECODING

We first provide an upper bound on the complexity exponent that guarantees a certain MAC-DMT performance $d(r)$.

Theorem 1: For the K-user MAC, the minimum over all ML-based decoders (all SD implementations, all halting and all decoding order policies) complexity exponent $c_{\mathrm{mac},d}(r)$ required to achieve a certain DMT $d(r) \le d^*_{\mathrm{mac}}(r)$, is upper bounded by

$$\bar{c}_{\mathrm{mac},d}(r) = \begin{cases} \sup_{\mu \in \mathcal{B}(r)} (K - n_r) r T \\ \quad + T \sum_{i=1}^{\nu} \left(r - (1 - \mu_i)^+ \right)^+, \text{ if } K > n_r, \\ \\ \sup_{\mu \in \mathcal{B}(r)} T \sum_{i=1}^{\nu} \left[\min\{r, r + \mu_i - 1\} \right]^+, \\ \quad\quad\quad\quad\quad\quad\quad\quad\quad \text{if } K \le n_r, \end{cases} \tag{8}$$

where $\mu = [\mu_1 \ \cdots \mu_\nu]^\top$, $\nu = \min\{K, n_r\}$ and

$$\mathcal{B}(r) := \left\{ \mu : \begin{array}{l} \mu_1 \ge \cdots \ge \mu_\nu, 0 \le \mu_i \in \mathbb{R} \\ \sum_{i=1}^{\nu} \left(|K - n_r| + 2i - 1 \right) \mu_i \le d(r) \end{array} \right\}.$$

Proof: See Appendix A. ∎

Remark 1: A remark is in order, regarding the high complexity present in the under-determined MAC case of $n_r < K$, and particularly regarding the term $(K - n_r) r T$ appearing in (8) (and the same in (9) appearing later for $T = 1$), which results in an increased complexity exponent, irrespective of the desired diversity performance. To gain some insight on this, we quickly recall that a sphere decoder performs a QR decomposition of matrix M in (7), which though for $n_r < K$, results in a matrix R that is an upper trapezoid matrix, whose bottom row contains $2T(K - n_r) + 1$ nonzero entries. Therefore, prior to processing the root node of a sphere-decoding tree, the sphere decoder must first search exhaustively among $N^{2T(K - n_r)}$ combinations of the N-ary

PAM constellation points of the integer entries of s (cf. (7)). For our particular case of having single antenna transmitters, N suffices to be chosen as $N = \text{SNR}^{\frac{r}{2}}$, which explains the term $(K - n_r)rT$ in (8).

A. Complexity cost of achieving the optimal MAC-DMT

We here seek to derive the complexity exponent needed to achieve the MAC-DMT optimal $d^*_{\text{mac}}(r)$ (cf. (4)). To answer this question, we will need to understand the coding duration requirement T for such DMT optimality. Towards this, we recall that a first step towards coding that achieves the optimal DMT performance, was taken in [5], which showed that for the lightly-loaded multiplexing-gain region $r \leq \frac{n_r}{K+1}$, uncoded (QAM) transmission ($T = 1$) is in fact DMT optimal. The question of deriving codes that can achieve DMT optimality for all r, was resolved in [15]–[17], which provided — depending on the values of K and n_r — different DMT optimal lattice coding schemes by encoding over time, thus requiring $T > 1$.

The following theorem plays a crucial role in tightening the complexity exponent, by proving that uncoded transmission ($T = 1$, QAM) is indeed MAC-DMT optimal for all multiplexing gains. The following result holds for all isotropic channel probability distributions, and it extends the aforementioned result in [15]–[17], [27] to all r, all n_r and all K.

Theorem 2: Uncoded QAM for each user achieves DMT optimality in the K-user symmetric SIMO MAC for all r, K, n_r.

  Proof: See Appendix B. ∎

We can now apply Theorem 1 to provide upper bounds on the complexity exponent $c^*_{\text{mac}}(r)$ that guarantees the DMT optimal $d^*_{\text{mac}}(r)$. The bound is constructive and is optimized over all known ML-based decoders, and over all existing codes. The result holds for i.i.d. Rayleigh fading statistics and for any decoding order policy. It corresponds to a DMT optimal transceiver that employs QAM transmission with $T = 1$, a search radius $\delta > \sqrt{d^*_{\text{mac}}(r) \log \text{SNR}}$, and a decoding halting policy that naturally halts decoding at $\text{SNR}^{c^*_{\text{mac}}(r)}$ flops.

Theorem 3: For the K-user MAC, the minimum over all codes and all ML-based decoders (all SD implementations, all halting and all decoding order policies) complexity exponent $c^*_{\text{mac}}(r)$ required to achieve the optimal DMT $d^*_{\text{mac}}(r)$, is upper bounded by

$$\bar{c}_{\text{mac}}(r) = \begin{cases} \sup_{\boldsymbol{\mu} \in \mathcal{B}(r)} (K - n_r)r + \sum_{i=1}^{\nu} \left(r - (1 - \mu_i)^+ \right)^+, \\ \qquad\qquad\qquad\qquad \text{if } K > n_r \\ \sup_{\boldsymbol{\mu} \in \mathcal{B}(r)} \sum_{i=1}^{\nu} \left[\min\{r, r + \mu_i - 1\} \right]^+, \\ \qquad\qquad\qquad\qquad \text{if } K \leq n_r, \end{cases} \quad (9)$$

where $\boldsymbol{\mu} = [\mu_1 \cdots \mu_\nu]^\top$, $\nu = \min\{K, n_r\}$ and

$$\mathcal{B}(r) := \left\{ \boldsymbol{\mu} : \begin{array}{l} \mu_1 \geq \cdots \geq \mu_\nu, 0 \leq \mu_i \in \mathbb{R} \\ \sum_{i=1}^{\nu} (|K - n_r| + 2i - 1) \mu_i \leq d^*_{\text{mac}}(r) \end{array} \right\}.$$

  Proof: It is a direct consequence of Theorems 1 and 2. ∎

The following corollaries provide explicit expressions for $\bar{c}_{\text{mac}}(r)$ for the single-input single-output (SISO) case of $n_r = 1$, and for $n_r = K$.

Corollary 4: For the K-user SISO MAC ($n_r = 1$), the complexity exponent required to achieve the optimal $d^*_{\text{mac}}(r)$, is upper bounded by

$$\bar{c}_{\text{mac}}(r) = (K - 1)r, \quad \text{for } 0 \leq r \leq \frac{1}{K}. \quad (10)$$

Corollary 5: For the K-user MAC with $n_r = K$, the complexity exponent required to achieve the optimal $d^*_{\text{mac}}(r)$, is upper bounded by

$$\bar{c}_{\text{mac}}(r) = r \left\lfloor \sqrt{K(1-r)} \right\rfloor$$
$$+ \left(r - 1 + \frac{K(1-r) - (\lfloor \sqrt{K(1-r)} \rfloor)^2}{2 \lfloor \sqrt{K(1-r)} \rfloor + 1} \right)^+. \quad (11)$$

Fig. 1 plots $\bar{c}_{\text{mac}}(r)$ for $K = 4, 5$ and $n_r = 1$, while Fig. 2 plots $\bar{c}_{\text{mac}}(r)$ for $n_r = K = 3, 4, 5$. In interpreting the plots below, we recall that the here maximum achievable multiplexing gain is $\frac{n_r}{K}$.

Remark 2: As we see from (11) and Fig. 2, when $n_r = K$, the complexity exponent $c(r)$ is not monotonically increasing. This is simply because an increasing r might indeed increase the objective function in the maximization, but at the same time it also decreases $d(r)$ and thus also decreases the volume of $\mathcal{B}(r)$ over which maximization takes place.
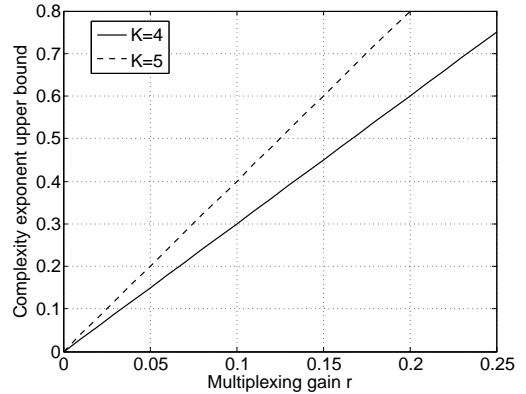

Fig. 1. Bound on the complexity exponent guaranteeing DMT optimality in the K-user SISO MAC: $K = 4, 5$ and $n_r = 1$.

B. The benefit of increasing the number of receive antennas

In the following we show the rather surprising result that a substantial increase in the number n_r of receive antennas, allows for maximal reductions in the complexity exponent. This is indeed surprising because, as n_r increases, a MAC-DMT optimal decoder must handle channels that are increasingly more singular [4] and thus typically harder to decode because near-singular channels typically result in denser signaling constellations at the receiver.

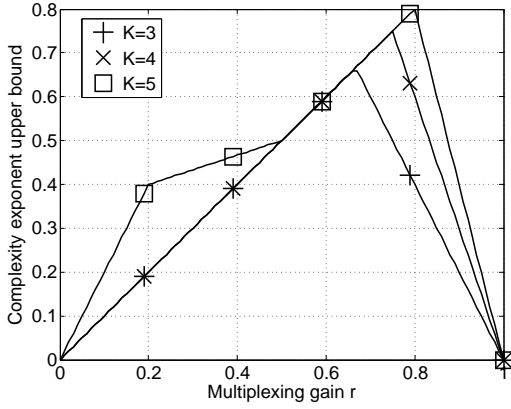[4] Recall from (4) that for $n_r > K$ the optimal MAC DMT is $d^*_{\text{mac}}(r) = n_r(1 - r)^+$.

Fig. 2. Bound on the complexity exponent guaranteeing DMT optimality in the K-user SIMO MAC: $K = 3, 4, 5$ and $n_r = K$.

Corollary 6: For the K-user SIMO MAC with $n_r \gg K$, the complexity exponent for achieving the optimal MAC-DMT $d^*_{\text{mac}}(r)$, approaches

$$\bar{c}_{\text{mac}}(r) = \min\left\{r, \frac{K-1}{n_r - K + 1}(1 - r)\right\}$$

and in the limit of asymptotically large n_r, the complexity exponent tends to zero.

Proof: See Appendix C. ∎

In Fig. 3(a) we plot the complexity-exponent upper bounds $\bar{c}_{\text{mac}}(r)$ for $K = 3$ users and for various values of n_r. The corresponding optimal diversity gains $d^*_{\text{mac}}(r)$ are shown in Fig. 3(b). It can be easily seen that the bounds $\bar{c}_{\text{mac}}(r)$ (with r fixed) decrease monotonically for an increasing n_r, and in the limit of very large n_r, the complexity exponent approaches 0, meaning that the decoder $\mathcal{D}_r$ — with the proper halting policy — can deliver the substantial diversity performance $n_r(1-r)$ in a computationally inexpensive manner.

III. PERFORMANCE AND COMPLEXITY FOR MAC ML-BASED DECODING WITH FEEDBACK-AIDED USER SELECTION

We now explore the ramifications of feedback-aided *user-selection* on the performance and complexity of ML-based MAC communications. Our motivation in exploring user selection comes from the associated gains in MAC reliability which — particularly for the high multiplexing-gain region — is typically quite low. This same method simultaneously allows for a "simplification" of the communication problem, from a larger multiple access channel, to a selectively reduced smaller and more manageable setting. Finally, this ability to provide simpler and more reliable communications, comes at a very reasonable feedback cost, which renders user selection applicable in the presence of limited feedback links.

We will here consider a MAC user-selection scheme, which is modified from the Jiang and Varanasi (JV) antenna selection algorithm in [25]. The JV algorithm, to the best of our knowledge, is currently the best performing antenna-selection algorithm for the point-to-point (single-pair) MIMO scenario.
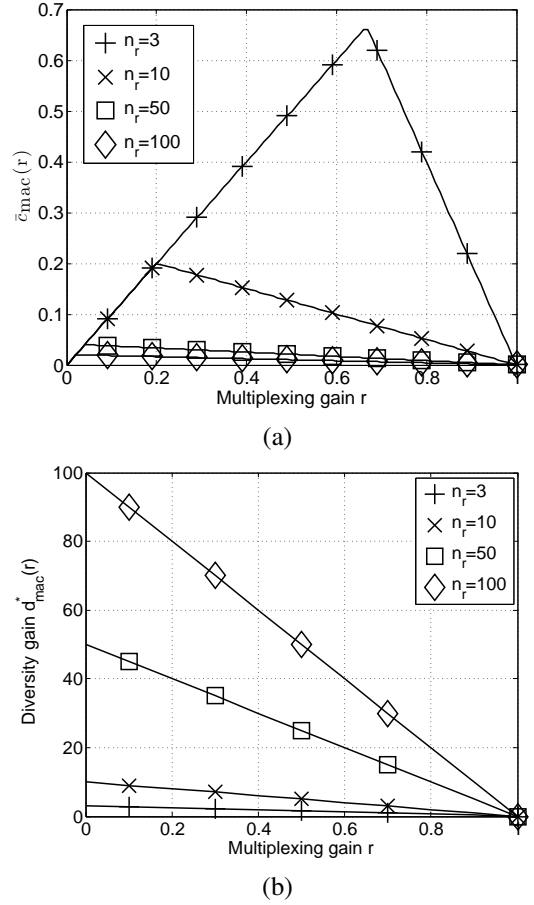

(a)

(b)

Fig. 3. Complexity-exponent upper bound $\bar{c}_{\text{mac}}(r)$ (see subfigure (a)) for achieving $d^*_{\text{mac}}(r)$ (see subfigure (b)) for $K = 3$ and for increasing n_r.

The proposed user selection scheme, to be described in more detail in Appendix D, selects which L out of K users will transmit throughout the coherence period of the channel, and then informs all users of the selection outcome via a feedback channel using $\log_2 \binom{K}{L}$ bits per channel coherence period. This selection decision is taken as a function of the entries of a matrix $\boldsymbol{R}_{\text{JV}}$ derived from the QR-decomposition $\boldsymbol{H}_{\text{eq}}\boldsymbol{\Pi} = \mathbf{Q}_{\text{JV}}\boldsymbol{R}_{\text{JV}}$ where $\boldsymbol{\Pi}$ is a permutation matrix, where $\mathbf{Q}_{\text{JV}}$ is a unitary matrix, and where — depending on the values of K, L, and n_r — $\boldsymbol{R}_{\text{JV}}$ can be either upper triangular or upper trapezoidal. This distinction in the shape of $\boldsymbol{R}_{\text{JV}}$ will here limit our consideration of L to an allowable set

$$\mathcal{L} := \{1, 2, \cdots, \nu, K\}$$

where we recall that $\nu = \min\{K, n_r\}$, and where $L = K$ corresponds to the case where no users are pruned out and thus where all K users transmit simultaneously to the receiver. Such choice of $L = K$ can, especially at high values of multiplexing gains, provide for the highest reliability.

We proceed to bound the DMT gains of user selection, and — based on these bounds — to then show that any such gains can come with a *simultaneous* exponentially-reduced complexity, compared to the complexity associated to $c^*_{\text{mac}}(r)$ in Theorem 3.

Remark 3: We note that under the adopted assumption that

the fading statistics are uniform across users, then the user selection algorithm will in fact 'prune all users (statistically) evenly and thus 'fairly. It is important to note that the rate considered here, does indeed account for the periods of 'silence of a user that has not been selected. In other words, the derived diversity-multiplexing expressions — which are the same for all users — correspond to a user rate that is simply the total number of bits sent by each user, divided by the total number of time-slots including the time-slots during which the user was kept silent by the selection algorithm.

A. Selection-aided MAC DMT bounds, for a fixed L

We first proceed with an upper bound on the selection-aided DMT $d_{\text{us},L}(r)$ for the employed user selection scheme, when the number of selected users is fixed to a certain L.

Theorem 7: Given L, and given the JV-inspired user selection scheme, the MAC DMT is upper bounded as

$$d_{\text{us},L}(r) \leq \bar{d}_{\text{us},L}(r) := \min_{\substack{k \geq 0, \ell \geq 1 \\ k+\ell \leq L}} d_{k,\ell}\left(\frac{\ell K r}{L}\right) \quad (12)$$

where

$$d_{k,\ell}(r) := \inf_{\mathcal{A}_\ell(r)} D_{k,\ell}(\boldsymbol{\alpha}),$$

$$\mathcal{A}_\ell(r) := \left\{ 0 \leq \alpha_1 \leq \cdots \leq \alpha_\ell : \sum_{i=1}^{\ell} (1-\alpha_i)^+ \leq r \right\}$$

and

$$D_{k,\ell}(\boldsymbol{\alpha}) := \sum_{i=1}^{\ell} (n_r + \ell - 2i + 1)\alpha_i + \sum_{i=1}^{\ell-1} (K - k - \ell)\alpha_i \\ + \alpha_\ell (K - k - \ell)(n_r - k - \ell + 1). \quad (13)$$

Proof: See Appendix E. ∎

We here note that it is not hard to show when $L = K$, that

$$d_{\text{us},K}(r) = d_{\text{mac}}^*(r) = \bar{d}_{\text{us},K}(r). \quad (14)$$

Also when $n_r = 1$, it is easy to show that

$$d_{\text{us},1}(r) = \bar{d}_{\text{us},1}(r). \quad (15)$$

Hence the bound is tight for $L = 1$ and $L = K$.

In Fig. 4 we plot the DMT upper bounds $\bar{d}_{\text{us},L}(r)$ from Theorem 7, for the (under-determined) case of $K = 4, n_r = 3$, with $L = 1, 2, 3$. In the same figure we compare the above $\bar{d}_{\text{us},L}(r)$ to the optimal MAC DMT $d_{\text{mac}}^*(r)$ corresponding to having no selection (or equivalently to having $L = K = 4$). Fig. 5 plots the DMT bounds for the (over-determined) case of $K = 3, n_r = 4$, with $L = 1, 2, 3$. In interpreting the figures, one must recall that, due to the fact that each user is selected with probability $\frac{L}{K}$, maintaining an average multiplexing gain r, will require that each transmitting user communicates at a multiplexing gain of $\frac{K}{L}r$. This is indeed taken into account, and the multiplexing gain reflects the true rate of communication, in full consideration of the fact that part of the time, some users are not transmitting. It should also be noted that for $L \leq \nu$, the maximal achievable multiplexing gain is $\frac{L}{K}$ or each user. Furthermore, it can be seen from the curves in Fig. 4 and

Fig. 5 that the DMT performance of proposed feedback-aided user-selection scheme strongly depends on the choice of L for different regions of multiplexing gain values r. For instance, in Fig. 5 of $K = 3$ and $n_r = 4$, choosing $L = 1$ yields the largest diversity gain whenever $r \in (0, 0.2]$. For $r \in (0.2, 0.5]$, $L = 2$ becomes the best choice, and if the desired multiplexing gain $r \in (0.5, 1]$, one should set $L = K = 3$. A similar observation can be made in Fig. 4. These cross-points between curves $\bar{d}_{\text{us},L}(r)$ for various L suggest further diversity gains by optimizing over L as a function of r. This will be discussed in more detail in Section III-C.

Another interesting observation from Fig. 4 and Fig. 5 is that despite the different configurations of K and n_r in these figures, the respective initial values $\bar{d}_{\text{us},L}(0)$ are the same for $L = 1, 2, 3$. This is actually no coincidence. To see it, note that as multiplexing gain r approaches zero from the right, meaning that the desired transmission rate R is arbitrarily close to zero, the only way for the communication system being in outage is exactly when all the *major* channel links are in deep-fade, i.e., when $\alpha_1 = \ldots = \alpha_\ell = 1$ in the set $\mathcal{A}_\ell(r)$. In this case, the function $D_{k,\ell}(\boldsymbol{\alpha} = \mathbf{1})$ simplifies to

$$D_{k,\ell}(\mathbf{1}) = K n_r - k(K + n_r) + k^2 + k\ell,$$

hence the minimal $D_{k,\ell}(\mathbf{1})$ occurs at $\ell = 1$ and

$$\bar{d}_{\text{us},L}(0) = \min_{0 \leq k \leq L-1} K n_r - k(K + n_r) + k^2 + k.$$

This shows that the value of $\bar{d}_{\text{us},L}(0)$ is symmetric between K and n_r whenever $L \leq \min\{K, n_r\}$. Moreover, with $K = 4$ and $n_r = 3$ (and the same holds for $K = 3$ and $n_r = 4$) we have $\bar{d}_{\text{us},L}(0) = \min_{0 \leq k \leq L-1} k^2 - 6k + 12$, and it is straightforward to see that $\bar{d}_{\text{us},1}(0) = 12$, $\bar{d}_{\text{us},2}(0) = 7$, $\bar{d}_{\text{us},3}(0) = 4$, and $\bar{d}_{\text{us},4}(0) = 3$ as shown in Fig. 4.
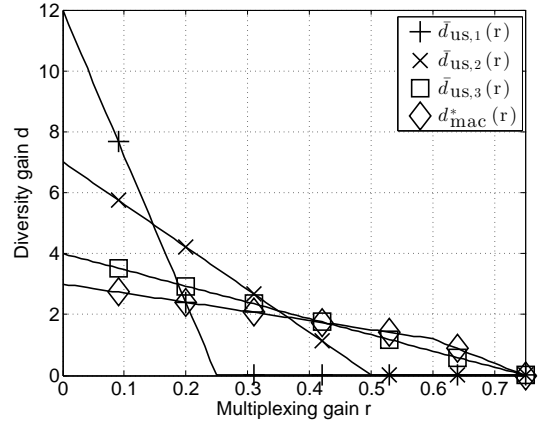

Fig. 4. Selection-aided DMT upper bounds $\bar{d}_{\text{us},L}(r)$, vs. the optimal MAC DMT $d_{\text{mac}}^*(r)$ without selection. $K = 4$, $n_r = 3$, and $L = 1, 2, 3$.

B. Selection-aided MAC complexity bounds, for a fixed L

We now use the previous DMT bound to upper bound the complexity exponent $c_{\text{us},L}(r)$ required to achieve this selection-aided DMT $d_{\text{us},L}(r)$ corresponding to the proposed L-user selection algorithm, for any fixed $L \in \mathcal{L}$. The result
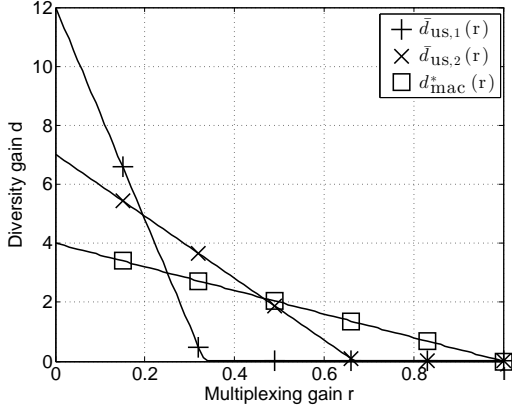
Fig. 5. Selection-aided DMT upper bounds $\bar{d}_{\text{us},L}(r)$, vs. the optimal MAC DMT $d_{\text{mac}}^*(r)$ without selection. $K = 3$, $n_r = 4$, and $L = 1, 2$.

is a direct consequence of Theorem 3, and it holds for i.i.d. Rayleigh fading channel statistics.

Theorem 8: For the K-user MAC, the complexity exponent $c_{\text{us},L}(r)$ required to achieve the selection-aided DMT $d_{\text{us},L}(r)$, for a given $L \leq \nu$, is upper bounded as

$$c_{\text{us},L}(r) \leq \bar{c}_{\text{us},L}(r) = \sup_{\boldsymbol{\alpha} \in \mathcal{F}(r)} \sum_{i=1}^{L} \left[\min\left\{ \frac{K}{L}r, \frac{K}{L}r + \alpha_i - 1 \right\} \right]^+ \tag{16}$$

where

$$\mathcal{F}(r) := \left\{ \boldsymbol{\alpha} : \begin{array}{l} \alpha_1 \leq \cdots \leq \alpha_L, 0 \leq \alpha_i \in \mathbb{R} \\ D_{0,L}(\boldsymbol{\alpha}) \leq \bar{d}_{\text{us},L}(r) \end{array} \right\}$$

and where $D_{k,\ell}(\boldsymbol{\alpha})$ is defined in (13). The performance-complexity pair $(d_{\text{us},L}(r), c_{\text{us},L}(r))$ is achieved with uncoded QAM ($T = 1$), a sphere decoder with a search radius $\delta > \sqrt{\bar{d}_{\text{us},L}(r) \log \text{SNR}}$, any decoding ordering, and a decoding halting policy that halts decoding if $N_{\max}(r) \doteq \text{SNR}^{c_{\text{us},L}(r)}$.

Proof: The result holds directly from Theorem 3 and from the fact that the complexity needed to achieve $d_{\text{us},L}(r)$ is upper bounded by the complexity needed to achieve $\bar{d}_{\text{us},L}(r) \geq d_{\text{us},L}(r)$. ■

C. Bound-based performance and complexity optimization over L

The performance and complexity bounds in Theorems 7 and 8 suggest gains by optimizing over L as a function of r. Using the bounds as an indicator to the best choice of L, we let

$$L^*(r) = \arg \max_{L \in \mathcal{L}} \{\bar{d}_{\text{us},L}(r)\}, \tag{17}$$

where we recall that $\mathcal{L} = \{1, 2, \ldots, \nu, K\}$. Hence

$$\bar{d}_{\text{us},L^*}(r) = \max_{L \in \mathcal{L}} \{\bar{d}_{\text{us},L}(r)\} \tag{18}$$

can serve as an upper bound on the selection-aided DMT $\max_{L \in \mathcal{L}} \{d_{\text{us},L}(r)\}$ maximized — at any given r — over the choices of L, including over the choice $L = K$ that all users are selected. As a result, going back to (16), we can calculate

$\bar{c}_{\text{us},L^*}(r)$ as a bound for the complexity required to achieve the optimized selection-aided DMT $\max_{L \in \mathcal{L}} \{d_{\text{us},L}(r)\}$.

For the under-determined case of $K = 4, n_r = 3$, Fig. 6 plots $L^*(r)$, while Fig. 7 plots the optimized, over L, selection-aided DMT upper bound $\bar{d}_{\text{us},L^*}(r)$. This bound is also compared to the bound $\bar{d}_{\text{us},L^{**}}(r)$ where

$$L^{**}(r) = \arg \max_{L \leq \nu} \{\bar{d}_{\text{us},L}(r)\} \tag{19}$$

corresponding to the case where selection always happens, i.e., corresponding to the case where we do not allow for $L = K$, even if $L^*(r) = K$.

Fig. 8 plots the complexity exponent upper bound $\bar{c}_{\text{us},L^*}(r)$ that guarantees the optimal — over all L — selection-aided DMT $d_{\text{us},L^*}(r)$. This complexity exponent bound is compared to $\bar{c}_{\text{mac}}(r)$ that achieves $d_{\text{mac}}^*(r)$ without user selection. The same figure also gives an interesting comparison to the complexity exponent $\bar{c}_{\text{us},L^{**}}(r)$ which, in conjunction with $\bar{d}_{\text{us},L^{**}}(r)$ in Fig. 7, suggests that the choice of $L^{**}(r)$ — rather than $L^*(r)$ — can provide substantial gains in complexity, with possibly moderate losses in performance.
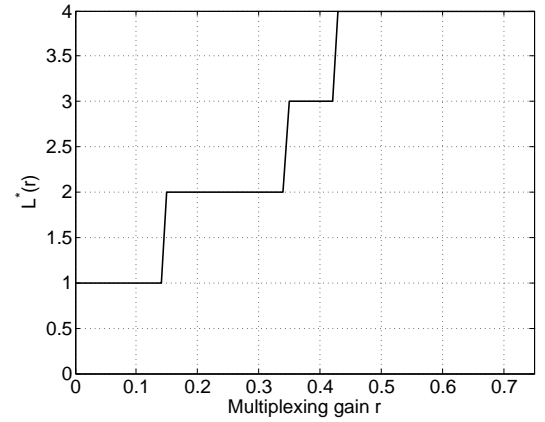

Fig. 6. Optimal number of selected users $L^*(r)$ for $K = 4$-user MAC with $n_r = 3$.
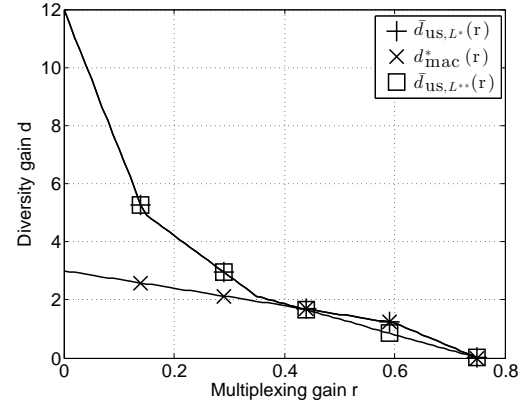

Fig. 7. DMT upper bounds $\bar{d}_{\text{us},L^*}(r)$ for user selection, compared to $d_{\text{mac}}^*(r)$ without selection, and $\bar{d}_{\text{us},L^{**}}(r)$ when user-selection is enforced, for the $K = 4$-user MAC with $n_r = 3$.
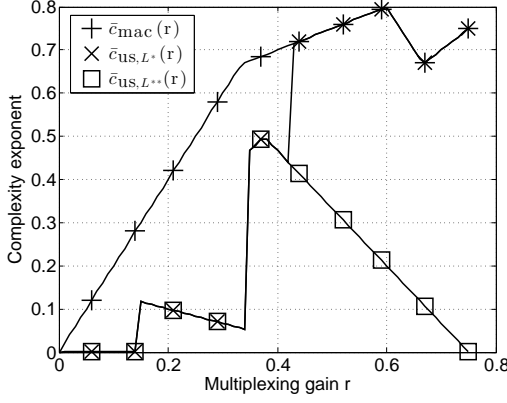
Fig. 8. Complexity exponent upper bounds $\bar{c}_{\text{us},L^*}(r)$ for user selection, compared to $\bar{c}_{\text{mac}}(r)$ without selection, and compared to $\bar{c}_{\text{us},L^{**}}(r)$ corresponding to when user-selection is enforced. Plots for $K = 4$ and $n_r = 3$.

D. Selection-aided complexity reduction, at no performance costs

We here explore the particular case where user selection is calibrated to reduce complexity, while maintaining reliability. The following quantifies this approach, based on the derived bounds. A crucial element in this effort is that any gains in the selection-aided DMT compared to the original MAC-DMT $d^*_{\text{mac}}(r)$, can allow us to effectively 'rest' the decoder often enough to reduce the complexity exponent, without falling below the required $d^*_{\text{mac}}(r)$. In this case, the resulting complexity exponent is upper bounded by

$$
\bar{c}_{\text{red-us},L}(r)
$$
$$
= \begin{cases} \sup\limits_{\boldsymbol{\alpha} \in \mathcal{F}_{\text{red-us},L}(r)} \sum_{i=1}^{L} \left[\min\left\{ \frac{K}{L}r, \frac{K}{L}r + \alpha_i - 1 \right\} \right]^+, \\ \qquad\qquad\qquad\qquad \text{if } \bar{d}_{\text{us},L}(r) \geq d^*_{\text{mac}}(r), \\ \\ \bar{c}_{\text{mac}}(r), \qquad\qquad\qquad\qquad \text{if otherwise,} \end{cases}
$$
(20)

where

$$
\mathcal{F}_{\text{red-us},L}(r) = \left\{ \boldsymbol{\alpha} : \begin{array}{l} \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_L, 0 \leq \alpha_i \in \mathbb{R} \\ D_{0,L}(\boldsymbol{\alpha}) \leq d^*_{\text{mac}}(r) \end{array} \right\}
$$

and where the second case in (20) refers to the situation where $L^* = K$. Hence an upper bound on the associated complexity exponent, minimized over L, is given by

$$
\bar{c}_{\text{red-us}}(r) = \min_{L \in \mathcal{L}} \bar{c}_{\text{red-us},L}(r).
$$
(21)

Fig. 9 explores this for the case of $K = 4$ and $n_r = 3$, presenting the resulting upper bounds $\bar{c}_{\text{red-us},L}(r)$ for all possible $L = 1, 2, 3$. The optimized complexity reduction, under the condition that we do not reduce reliability below $d^*_{\text{mac}}(r)$, is given in Fig. 10. The striking observation is that at low or moderate multiplexing gain regimes, the optimal $d^*_{\text{mac}}(r)$ can be achieved with a complexity exponent that can be as small as 0, i.e., with computational complexity that does not scale exponentially in the total number of codeword bits. At high multiplexing gains, no complexity reduction is available since $L^* = K$.
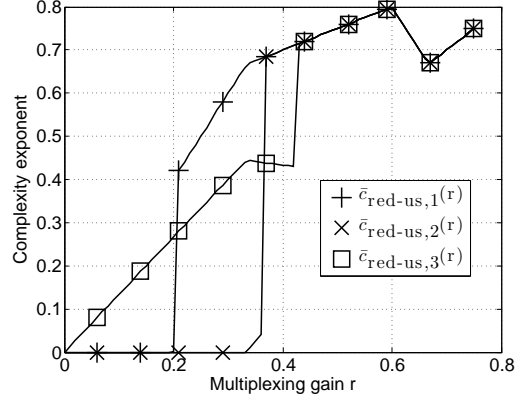

Fig. 9. Complexity exponent bounds $\bar{c}_{\text{red-us},L}(r)$ with $L = 1, 2, 3$, all guaranteeing $d^*_{\text{mac}}(r)$ performance for $K = 4$, and $n_r = 3$.
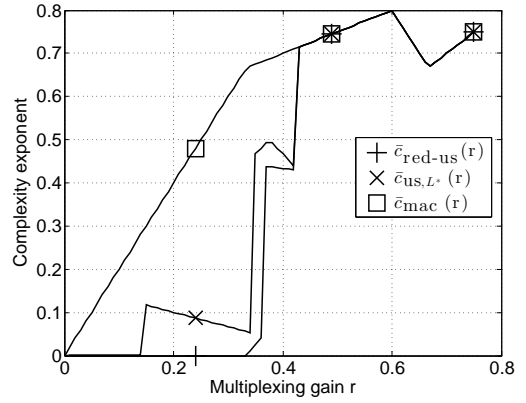

Fig. 10. Complexity exponent bounds $\bar{c}_{\text{red-us}}(r)$, $\bar{c}_{\text{us},L^*}(r)$ and $\bar{c}_{\text{mac}}(r)$, all guaranteeing $d^*_{\text{mac}}(r)$ performance for $K = 4, n_r = 3$.

IV. CONCLUSIONS

In this paper we derived bounds on the performance and complexity behavior of ML-based (and lattice-based) decoding, for the multiuser multiple access channel. Emphasis on ML-based decoding was motivated by the fact that MAC-related settings have distinctively reduced reliability, which hinders the use of low-performance decoders that would further reduce this reliability.

The derived complexity-vs-performance bounds were presented in the form of diversity and complexity exponents, and can provide insight on how to tradeoff performance and complexity in such outage-limited multiuser settings. The analysis shows that, under the requirement of efficient ML-based decoding, this tradeoff is not trivial; indeed we see that complexity constraints can crucially deteriorate performance.

The derived bounds also suggest substantial reliability and complexity benefits by increasing the number of receive antennas, as well as substantial benefits by utilizing just a few bits of feedback to allow for user selection. User selection is indeed pertinent as it can inherently increase reliability through channel selection, as well as can inherently decrease complexity by simply and selectively reducing the size of the problem at hand. These gains are highly sought in multiuser settings like

the MAC, which typically suffer from reduced reliability and increased complexity of decoding. In this context the analysis reveals the interesting finding that proper calibration of user selection can reduce the complexity exponent of near-optimal ML-based decoding, down to zero, thus revealing that — for a substantial range of multiplexing gains — the computational complexity of near-optimal ML-based decoding, need not scale exponentially in the total number of codeword bits.

APPENDIX A
PROOF OF THEOREM 1

In the following we establish an upper bound on the minimum complexity exponent $c_{\text{mac},d}(r)$ required by ML-based decoding to achieve a MAC DMT performance $d(r)$.

We recall from (7) that the sphere decoder 'sees' a channel model

$$\boldsymbol{y} = \boldsymbol{M}\boldsymbol{s} + \boldsymbol{w} \qquad (22)$$

where $\boldsymbol{M} = \text{SNR}^{\frac{1-r}{2}}\boldsymbol{H}\boldsymbol{G}$, where $\boldsymbol{H}$ is the equivalent channel in (6), and where $\boldsymbol{G}$ is the generator matrix of the overall (product) lattice code. To allow for sphere decoding in the under-determined ($K > n_r$), MMSE-preprocessing (cf. [26]) gives the MMSE-preprocessed code-channel matrix

$$\tilde{\boldsymbol{M}} = \left[\begin{array}{c} \boldsymbol{M} \\ \alpha\boldsymbol{I}_u \end{array} \right] = \boldsymbol{Q}\boldsymbol{R} \qquad (23)$$

where $\boldsymbol{I}_u$ is the $u \times u$ identity matrix, where $u = 2(K-n_r)T$, and where $\alpha = \text{SNR}^{-\frac{r}{2}}$. If $K \leq n_r$, then simply $\alpha = u = 0$. After the QR decomposition in (23), we get

$$\boldsymbol{r} := \left(\boldsymbol{R}^{\dagger} \right)^{-1} \boldsymbol{M}^{\dagger}\boldsymbol{y} = \boldsymbol{R}\boldsymbol{s} + \boldsymbol{w}'$$

where $\boldsymbol{w}' = -\alpha^2 \left(\boldsymbol{R}^{\dagger} \right)^{-1} \boldsymbol{s} + \left(\boldsymbol{R}^{\dagger} \right)^{-1} \boldsymbol{M}^{\dagger}\boldsymbol{w}$. For

$$\mathbb{S}_r := \left\{ \boldsymbol{s} \in \mathbb{Z}^{2KT} : \text{SNR}^{-\frac{r}{2}}\boldsymbol{G}\boldsymbol{s} \in \mathcal{X}_r \right\}$$

being the set of points in the product lattice Λ that constitute $\mathcal{X}_r$ after scaling, the sphere decoder takes the form

$$\hat{\boldsymbol{s}}_{\text{MMSE-SD}} = \arg\min_{\boldsymbol{s}\in\mathbb{S}_r} \|\boldsymbol{r} - \boldsymbol{R}\boldsymbol{s}\|^2 \qquad (24)$$

and it recursively enumerates all candidate vectors $\boldsymbol{s} \in \mathbb{S}_r$ within a given search sphere of radius $\delta = \sqrt{z \log \text{SNR}}$ for some $z > d(r)$.

To compute an upper bound on the complexity exponent, we follow the approach similar to [2]. Towards this, we first let

$$\lambda_i = \sigma_i(\boldsymbol{H}_{\text{eq}}^{\dagger}\boldsymbol{H}_{\text{eq}}), \quad i = 1, \ldots, \nu = \min\{K, n_r\}$$

be the nonzero singular values of $\boldsymbol{H}_{\text{eq}}^{\dagger}\boldsymbol{H}_{\text{eq}}$, arranged in ascending order, and then we let

$$\mu_i = -\frac{\log \lambda_i}{\log \text{SNR}}. \qquad (25)$$

Hence we can write

$$\sigma_i(\boldsymbol{R}) = \sigma_i(\tilde{\boldsymbol{M}}) = \sqrt{\alpha^2 + \sigma_i(\boldsymbol{M}^{\dagger}\boldsymbol{M})}, \quad i = 1, \ldots, 2KT. \qquad (26)$$

For $K \leq n_r$, we have

$$\sigma_i(\boldsymbol{R}) \doteq \text{SNR}^{\frac{1}{2}(1-r-\mu_{\lceil \frac{i}{2T} \rceil})}$$

and for $K > n_r$ we have

$$\sigma_i(\boldsymbol{R}) \doteq \left\{ \begin{array}{ll} \text{SNR}^{-\frac{r}{2}}, & \text{if } 1 \leq i \leq 2T(K-n_r) \\ \text{SNR}^{-\frac{r}{2}+\frac{1}{2}(1-\mu_j)^+}, & \text{otherwise} \end{array} \right\} \qquad (27)$$

where $j = \lceil \frac{i-2T(K-n_r)}{2T} \rceil$. In the above we have used that $\sigma_i(\boldsymbol{G}) \doteq \text{SNR}^0$, for all i.

We now see that for any given channel realization, corresponding to a specific $\boldsymbol{\mu} := [\mu_1 \cdots \mu_\nu]^\top$ (cf.(25)), the total number of visited nodes is given by

$$N_{\text{SD}}(\boldsymbol{\mu}) := \sum_{k=1}^{2KT} N_k(\boldsymbol{\mu})$$

where, drawing from [1, Lemma 1], we can see that

$$N_k(\boldsymbol{\mu}) \leq \prod_{i=1}^{k} \left[\sqrt{k} + 2\min\left\{ \frac{\delta}{\sigma_i(\boldsymbol{R})}, \sqrt{k}\text{SNR}^{\frac{r}{2}} \right\} \right]$$

and thus that

$$N_{\text{SD}}(\boldsymbol{\mu}) \leq \sum_{k=1}^{2KT} \prod_{i=1}^{k} \left[\sqrt{k} + 2\min\left\{ \frac{\delta}{\sigma_i(\boldsymbol{R})}, \sqrt{k}\text{SNR}^{\frac{r}{2}} \right\} \right].$$

Thus for the overdetermined case (cf. (26)), we have

$$N_{\text{SD}}(\boldsymbol{\mu}) \dot{\leq} \text{SNR}^{T\sum_{i=1}^{K}[\min\{r,r+\mu_i-1\}]^+} \qquad (28)$$

while for the underdetermined case (cf. (27)) we have

$$N_{\text{SD}}(\boldsymbol{\mu}) \dot{\leq} \text{SNR}^{(K-n_r)rT+T\sum_{i=1}^{n_r}\left(r-(1-\mu_i)^+\right)^+} \qquad (29)$$

where we have used the $2T$-fold multiplicity of the singular values of $\boldsymbol{H}$.

At this point, in the same spirit as in [2], the upper bound on the complexity exponent can be obtained as the solution to a constrained minimization problem of finding a value $c_{\text{mac},d}(r)$ such that the probability of a premature termination of SD algorithm is no larger than the channel outage probability, i.e.,

$$\Pr\left\{ N_{\text{SD}}(\boldsymbol{\mu}) \geq N_{\max}(r) = \text{SNR}^{c_{\text{mac},d}(r)} \right\} \leq \text{SNR}^{-d(r)}. \qquad (30)$$

This is the optimization reflected in the complexity exponent bound of Theorem 1. $\blacksquare$

APPENDIX B
PROOF OF THEOREM 2: OPTIMALITY OF UNCODED QAM

For $T = 1$, the rank-$2K$ overall (product) code-lattice Λ is isomorphic to the rectangular lattice $\mathbb{Z}^{2K}$, and the overall (product) code

$$\mathcal{X}_r = \left\{ \text{SNR}^{-\frac{r}{2}}\boldsymbol{x} : \boldsymbol{x} \in (\mathbb{Z}[\imath])^K, |x_i|^2 \leq \text{SNR}^r \right\}$$

is essentially uncoded (scaled) QAM. To show that the above code $\mathcal{X}_r$ satisfies the probabilistic complexity constraint (30), we follow the footsteps in [15], [16] and consider a K-fold extension of $\mathcal{X}_r$

$$\mathcal{X}_{r,\text{ext}} = \bigoplus_{i=1}^{K} \mathcal{X}_r \subset \text{SNR}^{-\frac{r}{2}} M_K(\mathbb{Z}[\imath])$$

where $M_K(\mathbb{Z}[\imath])$ denotes the set of $K \times K$ matrices with entries from $\mathbb{Z}[\imath]$. In other words, elements of $\mathcal{X}_{r,\text{ext}}$ are square $K \times K$ matrices whose entries are independent QAM constellation points after a scaling of $\text{SNR}^{-\frac{r}{2}}$. Naturally, for any given decoder, $\mathcal{X}_{r,\text{ext}}$ and $\mathcal{X}_r$ achieve the same DMT performance. Under brute force ML decoding, the error probability of $\mathcal{X}_{r,\text{ext}}$ is upper bounded by

$$
\begin{aligned}
&P_{e,\text{ext}}(r) \\
&= \mathbb{E}\Pr\left\{\boldsymbol{X}' \in \mathcal{E}(\boldsymbol{X}) \text{ decoded}\right\} \\
&= \mathbb{E}\Pr\left\{\bigcup_{k=1}^{K}\left\{\boldsymbol{X}' \in \mathcal{E}_k(\boldsymbol{X}) \text{ decoded}\right\}\right\} \\
&\overset{(i)}{\leq} \sum_{k=1}^{K} \mathbb{E}\Pr\left\{\boldsymbol{X}' \in \mathcal{E}_k(\boldsymbol{X}) \text{ decoded}\right\} \\
&\overset{(ii)}{\leq} \sum_{k=1}^{K} \mathbb{E}\Pr\left\{\bigcup_{\boldsymbol{X}' \in \mathcal{E}_k(\boldsymbol{X})}\left\{\boldsymbol{H}_{\text{eq}}: \right.\right. \\
&\qquad\qquad\qquad \left.\left.\left\|\text{SNR}^{\frac{1}{2}}\boldsymbol{H}_{\text{eq}}(\boldsymbol{X}-\boldsymbol{X}')\right\|^2 \overset{.}{\leq} 1\right\}\right\}
\end{aligned}
\tag{31}
$$

where the expectation is taken over all codeword matrices $\boldsymbol{X} \in \mathcal{X}_{r,\text{ext}}$, where $\mathcal{E}(\boldsymbol{X}) := \mathcal{X}_{r,\text{ext}} \setminus \{\boldsymbol{X}\}$ is the set of all possible erroneous decoded outputs given that codeword matrix $\boldsymbol{X}$ is transmitted, where $\mathcal{E}_k(\boldsymbol{X}) := \{\boldsymbol{X}' \in \mathcal{E}(\boldsymbol{X}) : \text{rank}(\boldsymbol{X}-\boldsymbol{X}') = k\}$ with $k = 1, \ldots, K$ forms a partition of $\mathcal{E}(\boldsymbol{X})$, where step (i) follows from the union bound, and where step (ii) is due to the use of a suboptimal bounded distance decoder (cf. [15]).

For any $\boldsymbol{X}' \in \mathcal{E}_k(\boldsymbol{X})$, set $\boldsymbol{\Delta}_{\boldsymbol{X}'} = \boldsymbol{X} - \boldsymbol{X}'$ and let $\boldsymbol{\Delta}_{\boldsymbol{X}'}\boldsymbol{\Delta}_{\boldsymbol{X}'}^{\dagger} = \boldsymbol{U}_{\boldsymbol{X}'}\boldsymbol{\Sigma}_{\boldsymbol{X}'}\boldsymbol{U}_{\boldsymbol{X}'}^{\dagger}$ be the corresponding eigen-decomposition. Note that as $\text{rank}(\boldsymbol{\Delta}_{\boldsymbol{X}'}) = k$, the eigenvalue matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}'}$ has form $\boldsymbol{\Sigma}_{\boldsymbol{X}'} = \text{diag}(\boldsymbol{\Omega}_{\boldsymbol{X}'}, \boldsymbol{0}_{K-k})$, where the diagonal of $\boldsymbol{\Omega}_{\boldsymbol{X}'}$ consists of all nonzero eigenvalues of $\boldsymbol{\Delta}_{\boldsymbol{X}'}\boldsymbol{\Delta}_{\boldsymbol{X}'}^{\dagger}$. Substituting the above into (31), we obtain

$$
\begin{aligned}
&\Pr\left\{\bigcup_{\boldsymbol{X}' \in \mathcal{E}_k(\boldsymbol{X})}\left\{\boldsymbol{H}_{\text{eq}}: \left\|\text{SNR}^{\frac{1}{2}}\boldsymbol{H}_{\text{eq}}\boldsymbol{\Delta}_{\boldsymbol{X}'}\right\|^2 \overset{.}{\leq} 1\right\}\right\} \\
&= \Pr\left\{\bigcup_{\boldsymbol{X}' \in \mathcal{E}_k(\boldsymbol{X})}\left\{\boldsymbol{G}_k: \text{SNR}\cdot\text{Tr}\left(\boldsymbol{G}_k^{\dagger}\boldsymbol{\Omega}_{\boldsymbol{X}'}\boldsymbol{G}_k\right) \overset{.}{\leq} 1\right\}\right\}
\end{aligned}
\tag{32}
$$

where $\boldsymbol{G}_k$ is an $(n_r \times k)$ random matrix with i.i.d. $\mathbb{CN}(0,1)$ entries. Let $\lambda_1 \leq \cdots \leq \lambda_m$ be the ordered nonzero eigenvalues of $\boldsymbol{G}_k^{\dagger}\boldsymbol{G}_k$, where $m = \min\{k, n_r\}$. Noting that $\det(\boldsymbol{\Omega}_{\boldsymbol{X}'}) \overset{.}{\geq} \text{SNR}^{-kr}$ and that $\text{Tr}(\boldsymbol{\Omega}_{\boldsymbol{X}'}) \overset{.}{\leq} 1$, it can be shown — drawing from [15] — that the condition of $\text{Tr}\left(\boldsymbol{G}_k^{\dagger}\boldsymbol{\Omega}_{\boldsymbol{X}'}\boldsymbol{G}_k\right) \overset{.}{\leq} 1$ implies that $\sum_{i=1}^{m}(1 - \mu_i)^+ \leq kr$, which is independent of the choice of $\boldsymbol{X}'$. Hence we have

$$
\begin{aligned}
&\Pr\left\{\bigcup_{\boldsymbol{X}' \in \mathcal{E}_k(\boldsymbol{X})}\left\{\boldsymbol{G}_k: \text{SNR}\cdot\text{Tr}\left(\boldsymbol{G}_k^{\dagger}\boldsymbol{\Omega}_{\boldsymbol{X}'}\boldsymbol{G}_k\right) \overset{.}{\leq} 1\right\}\right\} \\
&\overset{.}{\leq} \Pr\left\{\bigcup_{\boldsymbol{X}' \in \mathcal{E}_k(\boldsymbol{X})}\left\{\boldsymbol{\mu}: \sum_{i=1}^{m}(1 - \mu_i)^+ \leq kr\right\}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \Pr\left\{\boldsymbol{\mu}: \sum_{i=1}^{m}(1 - \mu_i)^+ \leq kr\right\} \\
&\overset{.}{=} \text{SNR}^{-d_{k,n_r}^*(kr)}
\end{aligned}
$$

where the last exponential equality follows from [6]. Finally, note that the error probability of $\mathcal{X}_r$, subject to (unterminated) joint ML decoding, is upper bounded by

$$
P_e(r) \leq \frac{1}{K}P_{e,\text{ext}}(r) \overset{.}{\leq} \frac{1}{K}\sum_{k=1}^{K}\text{SNR}^{-d_{k,n_r}^*(kr)} \overset{.}{=} \text{SNR}^{-d_{\text{mac}}^*(r)},
$$

which completes the proof. ∎

APPENDIX C
PROOF OF COROLLARY 6

First let us recall that for $n_r > K$, then $\nu = \min\{K, n_r\} = K$. Let us also recall from [6] that the joint probability density function of $\boldsymbol{\mu} = [\mu_1 \cdots \mu_K]^{\top}$ (cf. (25)) satisfies

$$
p(\boldsymbol{\mu}) \overset{.}{=} \text{SNR}^{-\sum_{i=1}^{K}(n_r-K+2i-1)\mu_i}
$$

provided that $\mu_i \geq 0$ for all i. As a result, for each $\boldsymbol{\mu} \in (\mathbb{R}^+)^K$, the corresponding complexity exponent is upper bounded by

$$
\bar{c}(\boldsymbol{\mu}) = \sum_{i=1}^{K}[\min\{r, r+\mu_i-1\}]^+
\tag{33}
$$

which is *not a function* of n_r. At this point, let us consider a decoder that decodes only when

$$
\sum_{i=1}^{K}(n_r - K + 2i - 1)\mu_i < d_{1,n_r}^*(r)
$$

while when $\sum_{i=1}^{K}(n_r - K + 2i - 1)\mu_i > d_{1,n_r}^*(r)$ the decoder simply declares an error. Since, for $n_r > K$, $d_{1,n_r}^*(r)$ dominates the MAC DMT (cf. [5]), the extra declared errors do not affect the overall diversity performance.

Now we note that for any $\mu_1 \geq \cdots \geq \mu_K \geq 0$ such that $\sum_i(n_r - K + 2i - 1)\mu_i = n_r(1 - r)$, we have that

$$
\sum_i(n_r - K + 1)\mu_i \leq n_r(1-r) \leq \sum_i(n_r - K + 2K - 1)\mu_i,
$$

and thus that

$$
\frac{n_r}{n_r + K - 1}(1-r) \leq \sum_i \mu_i \leq \frac{n_r}{n_r - K + 1}(1-r)
\tag{34}
$$

which, for $n_r \gg K$, implies that

$$
\frac{n_r}{n_r + K - 1}(1-r) \approx \frac{n_r}{n_r - K + 1}(1-r) \approx (1-r).
$$

Subject to the constraint in (34), we can see that the maximum value of $\bar{c}(\boldsymbol{\mu})$ in (33), is achieved by setting $\mu_1 = \frac{n_r}{n_r - K + 1}(1-r)$ and $\mu_2 = \cdots = \mu_K = 0$, where this maximal value takes the form

$$
\sum_{i=1}^{K}[\min\{r, r+\mu_i-1\}]^+ = \min\left\{r, \frac{K-1}{n_r - K + 1}(1-r)\right\}.
$$

Hence for $n_r \gg K$,

$$
\begin{aligned}
\bar{c}_{\text{mac}}(r) &= \sup_{\boldsymbol{\mu} \in \mathcal{B}(r)} \sum_{i=1}^{K} [\min\{r, r + \mu_i - 1\}]^+ \\
&\leq \sup_{\boldsymbol{\mu} \in \overline{\mathcal{B}}(r)} \sum_{i=1}^{K} [\min\{r, r + \mu_i - 1\}]^+ \quad (35) \\
&= \min\left\{ r, \frac{K-1}{n_r - K + 1}(1-r)^+ \right\}
\end{aligned}
$$

where

$$
\mathcal{B}(r) := \left\{ \boldsymbol{\mu} : \begin{array}{l} \mu_1 \geq \cdots \geq \mu_K, 0 \leq \mu_i \in \mathbb{R}, \\ \sum_{i=1}^{K} (K - n_r + 2i - 1)\mu_i \leq n_r(1-r)^+ \end{array} \right\}
$$

and

$$
\overline{\mathcal{B}}(r) := \left\{ \boldsymbol{\mu} : \begin{array}{l} \mu_1 \geq \cdots \geq \mu_K, 0 \leq \mu_i \in \mathbb{R} \\ \sum_{i=1}^{K} \mu_i \leq \frac{n_r}{n_r - K + 1}(1-r)^+ \end{array} \right\}.
$$

The inequality in (35) is due to the fact that $\mathcal{B}(r) \subseteq \overline{\mathcal{B}}(r)$, while note that equality holds when $\mu_1 = \frac{n_r}{n_r - K + 1}(1 - r)$, $\mu_2 = \cdots = \mu_K = 0$, which in turn implies that

$$
\bar{c}_{\text{mac}}(r) = \min\left\{ r, \frac{K-1}{n_r - K + 1}(1-r)^+ \right\}
$$

which establishes the corollary. ∎

APPENDIX D
DESCRIPTION OF USER-SELECTION SCHEME

The proposed user selection algorithm can be drawn from [25], by restricting selection to only transmit antennas (no selection of received antennas). In the setting with K single-antenna users, $\boldsymbol{H}_{\text{eq}}$ (cf. (6)) has as kth column the vector $\boldsymbol{h}_k$ corresponding to the channel vector of the kth user. The goal is to select L out of K users for transmission ($L \leq \nu = \min\{K, n_r\}$). The selection process is closely related to the QR-decomposition of $\boldsymbol{H}_{\text{eq}}$ using a Householder transformation, which exhibits good numerical stability. The selection takes L iterations, and each iteration consists of two steps, a matrix-column permutation followed by a Householder transformation. At the first iteration, the algorithm begins by finding the column $\boldsymbol{h}_{j_1}$ of $\boldsymbol{H}_{\text{eq}}$ with the largest column norm. Then it right-multiplies $\boldsymbol{H}_{\text{eq}}$ by a permutation matrix $\boldsymbol{\Pi}_1$ to swap $\boldsymbol{h}_1$ and $\boldsymbol{h}_{j_1}$. The second step is to apply a unitary Householder transformation $\boldsymbol{Q}_1$ to $\boldsymbol{H}_{\text{eq}}\boldsymbol{\Pi}_1$ such that the top-left entry of $\boldsymbol{Q}_1\boldsymbol{H}_{\text{eq}}\boldsymbol{\Pi}_1$ is the only nonzero (and positive) entry in the first column. After finishing with the first iteration, the algorithm shifts its focus to the trailing $(n_r - 1) \times (K - 1)$ submatrix $\boldsymbol{H}_1$ of $\boldsymbol{Q}_1\boldsymbol{H}_{\text{eq}}\boldsymbol{\Pi}_1$. Similar to the first iteration, the algorithm identifies the column with the largest norm in $\boldsymbol{H}_1$, swaps it with the first column, and then applies the Householder transformation. Thus, at the end of this iteration, the channel matrix becomes $\boldsymbol{Q}_2\boldsymbol{Q}_1\boldsymbol{H}_{\text{eq}}\boldsymbol{\Pi}_1\boldsymbol{\Pi}_2$. The third iteration is the same as the previous ones but focuses on the trailing $(n_r - 2) \times (K - 2)$ submatrix. The same process is repeated L times, resulting in an output matrix

$\boldsymbol{R}_{\text{JV}} = \boldsymbol{Q}_L \cdots \boldsymbol{Q}_1 \boldsymbol{H}_{\text{eq}} \boldsymbol{\Pi}_1 \cdots \boldsymbol{\Pi}_L$. The users associated to the first L columns of $\boldsymbol{R}_{\text{JV}}$ are the selected ones [5].

APPENDIX E
PROOF OF THEOREM 7

Recalling from the description in Appendix D of the selection algorithm, to select the L users, there are L iterations of applying column permutation matrix $\boldsymbol{\Pi}_i$ and Householder transformation $\boldsymbol{Q}_i$ to obtain an output matrix $\boldsymbol{R}_{\text{JV}}$, which is of the form

$$
\boldsymbol{R}_{\text{JV}} = \boldsymbol{Q}_L \cdots \boldsymbol{Q}_1 \boldsymbol{H}_{\text{eq}} \boldsymbol{\Pi}_1 \cdots \boldsymbol{\Pi}_L
$$
$$
= \begin{bmatrix} r_{1,1} & * & \cdots & * & * & \cdots & * \\ & r_{2,2} & \cdots & * & * & \cdots & * \\ & & \ddots & \vdots & \vdots & \vdots & \vdots \\ & & & r_{L,L} & * & \cdots & * \\ & & & & \vdots & \vdots & \vdots \\ & & & & * & \cdots & * \end{bmatrix}.
$$

Let u_i be the user associated with the ith column of $\boldsymbol{R}_{\text{JV}}$, and let $\{u_1, u_2, \ldots, u_L\}$ be the set of selected users. Since entries of $\boldsymbol{H}_{\text{eq}}$ are i.i.d. $\mathbb{CN}(0, 1)$, to meet an average multiplexing gain r for each user, the selected user has to transmit at a possibly larger multiplexing gain of $\frac{K}{L} r$. For the set of 'modified' outage events

$$
\mathcal{O}_{k,\ell} := \{\text{the sum-rate of users } u_{k+1}, \cdots, u_{k+\ell} \text{ is in outage}\} \tag{36}
$$

for all $k \geq 0$, $\ell \geq 1$, and $k + \ell \leq L$, we will see later on — in the process of the proof — that the outage event considered by Jiang and Varanasi [25, Theorem 4.1], is a special case of the above events corresponding to $k = 0$ and $\ell = L$, i.e., a special case of the outage event $\mathcal{O}_{0,L}$.

A. DMT Analysis for error event $\mathcal{O}_{k,\ell}$

To analyze the DMT for the error event $\mathcal{O}_{k,\ell}$ for any $k \geq 0$, $\ell \geq 1$, and $k + \ell \leq L$, let $\boldsymbol{R}_{k,\ell}$ be the matrix resulting from applying $(k + \ell)$ iterations of the Jiang-Varanasi algorithm to overall matrix $\boldsymbol{H}_{\text{eq}}$. We partition matrix $\boldsymbol{R}_{k,\ell}$ as follows:

$$
\begin{aligned}
\boldsymbol{R}_{k,\ell} &= \boldsymbol{Q}_{k+\ell} \cdots \boldsymbol{Q}_1 \boldsymbol{H}_{\text{eq}} \boldsymbol{\Pi}_1 \cdots \boldsymbol{\Pi}_{k+\ell} \\
&= \begin{bmatrix} \boldsymbol{R}_L & \boldsymbol{R}_{C,U} & \boldsymbol{R}_{R,U} \\ & \boldsymbol{R}_{C,B} & \boldsymbol{R}_{R,M} \\ & & \boldsymbol{R}_{R,B} \end{bmatrix} \quad (37)
\end{aligned}
$$

where

$$
\boldsymbol{R}_L = \begin{bmatrix} r_{1,1} & \cdots & r_{1,k} \\ & \ddots & \vdots \\ & & r_{k,k} \end{bmatrix}
$$

$$
\boldsymbol{R}_{C,U} = \begin{bmatrix} r_{1,k+1} & \cdots & r_{1,k+\ell} \\ \vdots & \vdots & \vdots \\ r_{k,k+1} & & r_{k,k+\ell} \end{bmatrix}
$$

[5]It is worth mentioning that if $K = n_r = L$, then $\boldsymbol{R}_{\text{JV}}$ is an upper triangular matrix, and thus $\boldsymbol{H}_{\text{eq}}\boldsymbol{\Pi}_1 \cdots \boldsymbol{\Pi}_L$ has a QR-decomposition equal to $(\boldsymbol{Q}_L \cdots \boldsymbol{Q}_1)^{\dagger} \boldsymbol{R}_{\text{JV}}$. Also, as the selection focuses only on a trailing submatrix of $\boldsymbol{Q}_{m-1} \cdots \boldsymbol{Q}_1 \boldsymbol{H}_{\text{eq}} \boldsymbol{\Pi}_1 \cdots \boldsymbol{Q}_{m-1}$ during the m-th iteration, one cannot say that the ith column of $\boldsymbol{R}_{\text{JV}}$ has the ith largest norm among all columns, for $i = 2, \cdots, L$.

$$\boldsymbol{R}_{C,B} = \begin{bmatrix} r_{k+1,k+1} & \cdots & r_{k+1,k+\ell} \\ & \ddots & \vdots \\ & & r_{k+\ell,k+\ell} \end{bmatrix}$$

$$\boldsymbol{R}_{R,U} = \begin{bmatrix} r_{1,k+\ell+1} & \cdots & r_{1,K} \\ \vdots & \ddots & \vdots \\ r_{k,k+\ell+1} & \cdots & r_{k,K} \end{bmatrix}$$

$$\boldsymbol{R}_{R,M} = \begin{bmatrix} r_{k+1,k+\ell+1} & \cdots & r_{k+1,K} \\ \vdots & \ddots & \vdots \\ r_{k+\ell,k+\ell+1} & \cdots & r_{k+\ell,K} \end{bmatrix}$$

$$\boldsymbol{R}_{R,B} = \begin{bmatrix} r_{k+\ell+1,k+\ell+1} & \cdots & r_{k+\ell+1,K} \\ \vdots & \ddots & \vdots \\ r_{n_r,k+\ell+1} & \cdots & r_{n_r,K} \end{bmatrix}$$

and where the entries satisfy

$$|r_{i,i}|^2 \;\geq\; \sum_{m=i}^{n_r} |r_{m,j}|^2, \tag{38}$$

for $i = 1, 2, \ldots, k+\ell$ and for $j = i+1, \ldots, K$.

For

$$\boldsymbol{R}_C := \begin{bmatrix} \boldsymbol{R}_{C,U} \\ \boldsymbol{R}_{C,B} \end{bmatrix}$$

the probability that the sum-rate of users $u_{k+1}, \cdots, u_{k+\ell}$ results in an outage, takes the form

$$\Pr\{\mathcal{O}_{k,\ell}\}$$
$$= \Pr\left\{ \log\det\left(\boldsymbol{I}_{n_r} + \mathrm{SNR}\boldsymbol{R}_C\boldsymbol{R}_C^\dagger\right) \;<\; \frac{K}{L}\ell r \log \mathrm{SNR} \right\}$$
$$= \int_{\mathcal{I}(r)} c \cdot \left[\prod_{i=1}^{k+\ell} p_{\chi^2_{2(n_r-i+1)}}\left(|r_{i,i}|^2\right) \left(\prod_{j=i+1}^{K} \frac{1}{\pi} e^{-|r_{i,j}|^2} \right) \right]$$
$$\times \frac{1}{\pi^{(K-k-\ell)(nr-k-\ell)}} e^{-\|\boldsymbol{R}_{R,B}\|^2} \, d\boldsymbol{R}_{k,\ell} \tag{39}$$
$$:\doteq \mathrm{SNR}^{-d_{k,\ell}\left(\frac{K}{L}\ell r\right)} \tag{40}$$

where c is a constant relating to ordered statistics (cf. the first constraint in (41)), where $p_{\chi^2_\kappa}(\cdot)$ is the probability density function for a χ^2 random variable with κ degrees of freedom and with mean $\frac{\kappa}{2}$, and where the integration region is

$$\mathcal{I}(r) := \left\{ \boldsymbol{R}_{k,\ell} : \begin{array}{c} |r_{i,i}|^2 \geq \sum_{m=i}^{n_r} |r_{m,j}|^2, \\ i = 1, \ldots, k+\ell, \quad j = i+1, \ldots, K \\ \det\left(\boldsymbol{I}_{n_r} + \mathrm{SNR}\boldsymbol{R}_C\boldsymbol{R}_C^\dagger\right) < \mathrm{SNR}^{\frac{K}{L}\ell r} \end{array} \right\}. \tag{41}$$

We have the following three remarks which we will jointly use later on.

Remark 4: We first note that the diversity exponent associated to $\Pr\{\mathcal{O}_{k,\ell}\}$ from (39), takes the form

$$d_{k,\ell}\left(\frac{K}{L}\ell r\right) = \lim_{\mathrm{SNR}\to\infty} -\frac{\log \Pr\{\mathcal{O}_{k,\ell}\}}{\log \mathrm{SNR}} = \inf_{\boldsymbol{R}_{k,\ell}\in\mathcal{I}(r)} D\left(\boldsymbol{R}_{k,\ell}\right)$$

where the last equality is a manifestation of Laplace's principle [6], and where $D\left(\boldsymbol{R}_{k,\ell}\right)$ is a specific diversity-exponent function corresponding to the integrand in (39). To understand the above, we shall look for sub-events of $\mathcal{I}(r)$ that yield the

smallest possible diversity exponent $D\left(\boldsymbol{R}_{k,\ell}\right)$. To this end, note that the entries in $\boldsymbol{R}_L$ and $\boldsymbol{R}_{R,U}$ are not involved in the second constraint in $\mathcal{I}(r)$, i.e., are not involved in the constraint $\det\left(\boldsymbol{I}_{n_r} + \mathrm{SNR}\boldsymbol{R}_C\boldsymbol{R}_C^\dagger\right) < \mathrm{SNR}^{\frac{K}{L}\ell r}$. Furthermore one can see that smaller absolute values of nonzero entries in $\boldsymbol{R}_L$ and $\boldsymbol{R}_{R,U}$ correspond to larger values of $D\left(\boldsymbol{R}_{k,\ell}\right)$. We hence conclude that the dominant sub-event of $\mathcal{I}(r)$, and the one yielding the smallest diversity exponent $D\left(\boldsymbol{R}_{k,\ell}\right)$), must consist of matrices $\boldsymbol{R}_{k,\ell}$ for which the nonzero entries of the associated submatrices $\boldsymbol{R}_L$ and $\boldsymbol{R}_{R,U}$ are the most typical ones. In other words, the smallest value of $D\left(\boldsymbol{R}_{k,\ell}\right)$ must result from the case when the nonzero entries of submatrices $\boldsymbol{R}_L$ and $\boldsymbol{R}_{R,U}$ have magnitudes in the order of SNR^0.

Remark 5: Let $\lambda_1 \geq \cdots \geq \lambda_\ell$ be the ordered singular values of $\boldsymbol{R}_C$, and let $\nu_1 \geq \cdots \geq \nu_\ell$ be the ordered singular values of $\boldsymbol{R}_{C,B}$. Clearly, as $\boldsymbol{R}_{C,B}\boldsymbol{R}_{C,B}^\dagger \preceq \boldsymbol{R}_C\boldsymbol{R}_C^\dagger$, we have

$$\nu_i^2 \leq \lambda_i^2, \quad i = 1, 2, \ldots, \ell. \tag{42}$$

Using [25, Lemma 3.3], we have

$$(\boldsymbol{R}_{C,B})_{i,i} = r_{k+i,k+i}^2 \geq \frac{\sum_{j=i}^{\ell}\nu_j^2}{\ell-i+1} \doteq \nu_i^2 \tag{43}$$

for $i = 1, 2, \ldots, \ell$. Moreover, from [25, Eq.(27)] we have that the squared diagonal elements in $\boldsymbol{R}_{C,B}$ are multiplicatively majorized by its squared singular values, i.e., that

$$\prod_{i=1}^{m}\nu_i^2 \geq \prod_{i=1}^{m} r_{k+i,k+i}^2 \dot{\geq} \prod_{i=1}^{m}\nu_i^2, \quad m = 1, 2, \ldots, \ell \tag{44}$$

where the second dotted inequality is due to (43). It then follows that

$$r_{k+i,k+i}^2 \doteq \nu_i^2 \tag{45}$$

for $i = 1, 2, \ldots, \ell$.

Remark 6: After the first k iterations of the Jiang-Varanasi algorithm, we get

$$\boldsymbol{Q}_k \cdots \boldsymbol{Q}_1 \boldsymbol{H}_{\mathrm{eq}} \boldsymbol{\Pi}_1 \cdots \boldsymbol{\Pi}_k$$
$$= \begin{bmatrix} r_{1,1} & * & \cdots & * & * & \cdots & * \\ & r_{2,2} & \cdots & * & * & \cdots & * \\ & & \ddots & \vdots & \vdots & \vdots & \vdots \\ & & & r_{k,k} & * & \cdots & * \\ & & & & \vdots & \vdots & \vdots \\ & & & & * & \cdots & * \end{bmatrix}$$
$$= \begin{bmatrix} \boldsymbol{R}_L & \boldsymbol{R}_R \\ \boldsymbol{0} & \end{bmatrix}$$

where $\boldsymbol{R}_R$ is the $(n_r \times (K-k))$ matrix consisting of the rightmost $(K-k)$ columns of the above matrix. In Remark 4 we argued that the dominant matrices $\boldsymbol{R}_{k,\ell}$ in event $\mathcal{I}(r)$ — yielding the smallest possible diversity exponent $D\left(\boldsymbol{R}_{k,\ell}\right)$ — must have submatrices $\boldsymbol{R}_L$ whose nonzero entries are of magnitude that is in the order of SNR^0. This implies that the entries of $\boldsymbol{R}_R$ can still be regarded as i.i.d. $\mathbb{CN}(0,1)$ random variables since any ensemble of random $\mathbb{CN}(0,1)$ variables has the same asymptotic probability as those subject to an additional constraint of magnitude lesser than SNR^0. Consequently, the ordered singular values $\lambda_i, i = 1, \ldots, \ell$,

of matrix $\boldsymbol{R}_C$ are still of the same joint probability density function as that for the ordered singular values of an $(n_r \times \ell)$ random matrix with i.i.d. $\mathbb{CN}(0,1)$ entries.

With the above three remarks in place, we now proceed to analyze the integral (39) to obtain a formula for the DMT function $d_{k,\ell}(r)$. Specifically, we will show that

$$d_{k,\ell}(r) = \inf_{\mathcal{A}_\ell(r)} \left\{ \sum_{i=1}^{\ell}(n_r+\ell-2i+1)\alpha_i + \sum_{i=1}^{\ell-1}(K-k-\ell)\alpha_i \right. $$
$$\left. + \alpha_\ell(K-k-\ell)(n_r-k-\ell+1) \right\} \quad (46)$$

where

$$\mathcal{A}_\ell(r) = \left\{ 0 \le \alpha_1 \le \cdots \le \alpha_\ell : \sum_{i=1}^{\ell}(1-\alpha_i)^+ \le r \right\}.$$

To see the above, set $\lambda_i^2 \doteq \text{SNR}^{-\alpha_i}$ with $\alpha_1 \le \alpha_2 \cdots \le \alpha_\ell$. The first summand appearing in (46) follows from the aforementioned fact that the joint probability density function of ordered singular values $\lambda_1 \ge \cdots \ge \lambda_\ell$ for an $(n_r \times \ell)$ matrix with i.i.d. $\mathbb{CN}(0,1)$ entries is (cf. [6], [25])

$$p(\alpha_1, \cdots, \alpha_\ell) \doteq \text{SNR}^{-\sum_{i=1}^{\ell}(n_r+\ell-2i+1)\alpha_i}.$$

Also by (38), (42), and (45), we have the following constraints for the entries in matrix $\boldsymbol{R}_{k,\ell}$:

1) For $i = 1, \ldots, \ell-1$, the entries $r_{k+i,j}$ ($j = k+\ell+1, \ldots, K$) must satisfy

$$|r_{k+i,j}|^2 \le |r_{k+i,k+i}|^2 \le \lambda_i^2.$$

The constraints on $r_{k+i,j}$ contribute to (46) the second summation, i.e. the term $\sum_{i=1}^{\ell-1}(K-k-\ell)\alpha_i$.

2) Entries $r_{k+i,j}$ with $i = \ell, \ldots, n_r-k$ and $j = k+\ell+1, \ldots, K$, must satisfy

$$\sum_{i=\ell}^{n_r-k}|r_{k+i,j}|^2 \le |r_{k+\ell,k+\ell}|^2 \doteq \mu_\ell^2 \le \lambda_\ell^2$$

Such constraints contribute to (46) the last summand, i.e., the term $\alpha_\ell(K-k-\ell)(n_r-k-\ell+1)$.

This completes the proof of (46).

Finally, the proof of Theorem 7 is complete after noting that the union of outage events $\mathcal{O}_{k,\ell}$, is a subset of the overall outage event, i.e., that

$$\bigcup_{\substack{k \ge 0, \ell \ge 1, \\ k+\ell \le L}} \mathcal{O}_{k,\ell} \subseteq \bigcup_{\mathcal{U} \subset \{u_1, \ldots, u_L\}} \{\boldsymbol{H}_{\text{eq}} : \text{users in } \mathcal{U} \text{ are in outage}\}.$$

∎

Remark 7: It is interesting to see that the antenna-selection DMT in Jiang and Varanasi ([25, Theorem 4.1]), can be derived as a special case of (46). Specifically it can be shown that when $k = 0$ and $\ell = L \le \nu$, the corresponding $d_{0,L}(r)$ coincides with the DMT in [25], as both functions are piecewise linear, connecting the following $(P+2)$ points

$$(r, (K-r)(n_r-r)), r = 0, 1, \ldots, P, \text{ and } (L, 0), \quad (47)$$

where

$$P = \arg\min_{p=0,1,\ldots,L-1} \frac{(K-p)(n_r-p)}{L-p}. \quad (48)$$

REFERENCES

[1] J. Jaldén and P. Elia, "Sphere decoding complexity exponent for decoding full-rate codes over the quasi-static MIMO channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5785–5803, Sep. 2012.

[2] A. Singh, P. Elia, and J. Jaldén, "Achieving a vanishing SNR gap to exact lattice decoding at a subexponential complexity," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3692–3707, Jun. 2012.

[3] A. Singh, P. Elia, , and J. Jaldén, "Complexity analysis for ML-based sphere decoder achieving a vanishing performance-gap to brute force ML decoding," in *Proc. Int. Zurich Seminar on Communications (IZS)*, Mar. 2012, pp. 127–130.

[4] P. Elia and J. Jaldén, "General DMT optimality of LR-aided linear MIMO-MAC transceivers with worst-case complexity at most linear in sum-rate," in *Proc. IEEE Information Theory Workshop (ITW)*, Jan. 2010.

[5] D. N. C. Tse, P. Viswanath, and L. Zheng, "Diversity-multiplexing tradeoff in multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1859–1874, Sept. 2004.

[6] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[7] G.R. Jithamithra and B. Sundar Rajan, "Minimizing the complexity of fast sphere decoding of STBCs," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6142–6153, Dec. 2013.

[8] L. P. Natarajan, K. Pavan Srinath and B. Sundar Rajan, "On the sphere decoding complexity of high rate multigroup decodable STBCs in asymmetric MIMO systems," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5959–5965, Sep. 2013.

[9] J. Jaldén and L.G. Barbero, B. Ottersten, and J.S. Thompson, "The error probability of the fixed-complexity sphere decoder," *IEEE Trans. Signal Processing*, vol. 57, no. 7, pp. 2711–2720, Jul. 2009.

[10] S. Sirinaunpiboon, Calderbank, A.R.; Howard, S.D. "Fast essentially maximum likelihood decoding of the Golden Code", *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3537–3541, Jun. 2011.

[11] L.P. Natarajan and B.S. Rajan, "An adaptive conditional zero-forcing decoder with full-diversity, least complexity and essentially-ML performance for STBCs," *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 253–263, Jan. 2013.

[12] A. Singh, P. Elia, and J. Jaldén, "Rate-reliability-complexity tradeoff for ML and lattice decoding of full-rate codes," in *Proc. 2013 IEEE Int. Symp. Inform. Theory*, ISIT 2013.

[13] J. Jaldén and P. Elia, "DMT optimality of LR-aided linear decoders for a general class of channels, lattice designs, and system models," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4765–4780, Sep. 2010.

[14] J. Jaldén and P. Elia, "LR-aided MMSE lattice decoding is DMT optimal for all approximately universal codes," in *Proc. 2009 IEEE Int. Symp. Inform. Theory*, ISIT 2009.

[15] H. F. Lu, C. Hollanti, R. Vehkalahti, and J. Lahtonen, "DMT optimal codes constructions for multiple-access MIMO channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3594–3617, June 2011.

[16] H. F. Lu, "Remarks on diversity-multiplexing tradeoffs for multiple-access and point-to-point MIMO channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 858–863, Feb. 2012.

[17] H.-F. Lu, R. Vehkalahti, C. Hollanti, J. Lahtonen, Y. Hong, and E. Viterbo, "New space-time code construcions for two user multiple access channels," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 6, pp. 939–957, Dec. 2009.

[18] S. Kittipiyakul, P. Elia, and T. Javidi, "High-SNR analysis of outage-limited communications with bursty and delay -limited information," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 746–763, Feb. 2009.

[19] T. Datta, A. Chockalingam, and E. Viterbo, "Gaussian sampling based lattice decoding," in *Proc. 2013 IEEE Int. Symp. Inform. Theory*, ISIT 2013.

[20] K. Singhal, T. Datta, and A. Chockalingam, "Lattice reduction aided detection in large-MIMO systems," *Proc. 2013 IEEE Int. workshop on Signal Processing Advances in Wireless Communications*, SPAWC 2013.

[21] J. Pan, Wing-Kin Ma, and J. Jaldén, "MIMO detection by Lagrangian dual maximum-likelihood relaxation: Reinterpreting regularized lattice decoding," *IEEE Trans. Signal Processing*, vol. 62, no. 2, pp. 511–524, Jan. 2014.

[22] P. Elia, F. Oggier, and P. V. Kumar, "Asymptotically optimal cooperative wireless networks with reduced signaling complexity," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 258–267, Feb. 2007.

[23] P. Elia and P. V. Kumar, "Approximately universal optimality over several dynamic and non-dynamic cooperative diversity schemes for wireless networks." Dec. 7, 2005 [Online]. Available: http://arxiv.org/pdf/cs.IT/0512028

[24] J. Jaldén and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1474-1484, Apr. 2005.

[25] Y. Jiang and M. K. Varanasi, "The RF-chain limited MIMO system-part I: optimum diversity-multiplexing tradeoff," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5238–5247, Oct. 2009.

[26] M. O. Damen, H. El Gamal, and G. Caire, "MMSE-GDFE lattice decoding for solving under-determined linear systems with integer unknowns," in *Proc. 2004 IEEE Int. Symp. Inform. Theory*, ISIT 2004.

[27] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, UK: Cambridge, University Press, 2005.